Analyseur syntaxique de questions en anglais : mode d'emploi

Objectifs :
Le but de cet analyseur est de reconnaître toutes les formes possibles de questions relativement simples en style direct pour l'entraînement autonome des étudiants.
Pour des raisons pragmatiques et pédagogiques je les ai réparties en 3 catégories : les questions existentielles et identitaires (portant sur l'existence ou l'identité du sujet par des prédications utilisant le verbe-copule BE) , les questions fermées (celles auxquelles on répond par « Oui » ou par « Non ») et les questions ouvertes (toutes les autres).
Les annotations syntaxiques sont focalisées sur les problèmes récurrents de maîtrise de la syntaxe:

a) distinction entre auxiliaires et verbes pleins,
b) distinction entre sujet et complément,
c) distinction entre constructions actives et passives,
d) combinaison des différentes formes aspectuelles.

L'analyseur est capable de détecter la plupart des erreurs usuelles des étudiants dans l'ordre des mots ou la morpho-syntaxe des verbes, mais aussi par rapport à la cohérence des nombres entre sujets, déterminants et forme verbales (« How many *is ») : l'automate sous-jacent est sensiblement plus complexe que ce que l'on utiliserait pour le traitement d'un corpus réputé syntaxiquement correct.

Prérequis :
Cet analyseur repose sur l'outil NooJ développé par Max Silberztein (pour plus d'information et pour télécharger cet outil voir le site de NooJ : http://www.nooj4nlp.net/pages/nooj.html
Pour utiliser cet analyseur de questions, il faut donc vous placer dans l'environnement NooJ (de préférence sur PC et avec un écran suffisamment grand ; il existe aussi une version Mac mais elle est plus compliquée à installer : il faut activer l'environnement X11 associé à la version MONO pour Mac-OS.) Même si vous ne connaissez pas NooJ, vous pouvez quand même apprendre très vite à utiliser l'outil : il vous suffit d'activer l'application NooJ et de :
1°) charger un fichier de texte à analyser par le menu : « File » → « Open » → « Text »
(NooJ peut convertir un fichier Word ou HTML vers son format interne .not)
2°) faire une première analyse lexicale (commande « Linguistic Analysis » du Menu « TEXT »)
3°) lancer l'analyseur de questions sur le texte (commande « Locate » du menu « TEXT ») ; vous obtiendrez en sortie une liste de concordances annotées contenant toutes les questions reconnues (que vous pourrez récupérer au format interne NooJ .noc ou exporter vers un traitement de texte).

Utilisation pour l'entraînement en auto-formation ou en classe :
Cet outil peut-être utilisé de deux façons :

- ***en mode DEBUG*** : charger l'analyseur par « File » → « Open » → « Grammar » et taper une question sur la ligne de saisie (les structures correctes s'affichent en vert, les formes avec des erreurs en rouge) ; pour l'utilisation en classe, on peut faire une projection et taper ou faire taper des questions produites par les étudiants et les faire réfléchir collectivement sur leurs erreurs ; dans une salle multimédia équipée de NooJ, chaque étudiant pourrait s'entraîner sur sa propre machine.
- ***en mode analyse de fichier et extraction de concordances :*** les étudiants peuvent traiter globalement un fichier (par exemple, toutes les réponses à une série d'exercices) puis éditer les concordances et réfléchir sur leurs erreurs à partir des annotations disponibles.

Limitations de l'analyseur :
L'outil ne traite que des phrases de complexité limitée (ni coordinations ni subordonnées, ni compléments multiples, pas non plus de prépositions rejetées en fin de phrase...) mais il est néanmoins suffisant pour un usage pédagogique réaliste.
NB1 : pour éliminer les résultats multiples produits par l'annotation lexicale des participes passés comme adjectifs j'ai « interdit » aux adjectifs attributs de se terminer en « ed » ; les « vrais » adjectifs en « ed » comme « naked », « rugged » ou « raven-haired » ne sont donc pas utilisables en position attribut (mais vous avez déjà de quoi faire avec tous les autres!).
NB2 : l'analyseur ignore la sémantique et acceptera des phrases bien formées n'ayant aucun sens.
NB3 : la cohérence syntaxique n'est garantie entre le sujet et la forme verbale que pour le genre et le nombre, les déterminants du sujet et la structure des compléments ne sont pas tous contrôlés.